PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 27 May 15, 1941 Number 5
Copyright 1941 by the National Academy of Sciences

*TWO INTERESTING NUCLEI OF PLANETARY NEBULAE: IC 418 AND NGC 40**

BY P. SWINGS AND O. STRUVE

McDONALD OBSERVATORY, UNIVERSITY OF TEXAS AND YERKES OBSERVATORY, UNIVERSITY OF CHICAGO

Communicated April 12, 1941

Attention has recently been directed by several investigators to the possible relation between Bethe's[1] process of energy generation and the chemical branching among the Wolf-Rayet shells and among the atmospheres of late-type stars. However the connection, if there is any, is not yet clear; hence it is important to collect additional observational material concerning the behavior of carbon and nitrogen in cosmic sources. Toward this end work was started at the McDonald Observatory in 1939. The program consisted of objects of high excitation: O stars (especially those of class Of, with vestiges of emission), Wolf-Rayet stars and planetary nuclei.

The subdivision of the Wolf-Rayet stars into a carbon sequence and a nitrogen sequence has been investigated in detail by C. S. Beals,[2] who has found that the split holds for all observed objects which are not surrounded by visible nebulosity. The subdivision may be readily extended to the absorption O stars with vestiges of emission. For example, in the case of 9 Sagittae and HD 192639, which resemble one another in their absorption spectra, the former shows strong N III and weak C III emission lines, whereas the opposite is true for the latter.

Until quite recently it had usually been assumed that the nuclei of planetary nebulae, when of the Wolf-Rayet type, also belonged to one of the two sequences. But actually, since most planetary nuclei are very faint, the assignment to one of the two branches was often doubtful. Among the objects on our observing program were NGC 6543 and Campbell's hydrogen envelope star (BD +30°3639). The nucleus of NGC 6543 had been tentatively classified as a *WN*6 star by Beals,[3] whereas C. H. Payne had placed it among the carbon stars.[4] Our spectrograms show conclusively that the characteristic lines of N IV, C IV and O IV appear

with similar intensities,[5] so that the object is intermediate between the *WN* and *WC* sequences. An investigation of IC 4997[6] showed that the nucleus of this stellar planetary also possesses lines of N III, C III and C IV and is, thus, of a type similar to NGC 6543 but of lower excitation. The investigation of BD 30°3639[7] called attention to the rather striking association of a pure carbon nucleus surrounded by a nebulosity rich in nitrogen. In the nucleus of BD 30°3639 there is no trace of nitrogen at any stage of ionization, whereas the red color of the nebular envelope is due to [N II] at least as much as to H_α. Another object, HD 167362, was found spectroscopically[8] to be exactly of the same type, although its distance is too great to allow the nebula to be resolved.

We have now investigated two other planetaries: IC 418, whose nucleus shows bright lines of N III, and C III with similar intensities, and NGC 40, which has a pure *WC* nucleus surrounded by a nebulosity rich in nitrogen.

IC 418.—The nucleus was tentatively classified as *WC7* by Beals,[3] whereas C. H. Payne[4] considered it as a star without carbon. Several spectrograms were secured in January, 1941, at the McDonald Observatory. We also had the privilege of using an excellent spectrogram taken by Dr. E. Hubble at the Mount Wilson Observatory. Nothing important could be added to the descriptions of the nebular spectrum by Wright[9] and by Stoy.[10] The red lines of [N II] are very strong, their combined intensity being similar to the intensity of H_α; [N II] 5755 is also fairly strong. But our spectrograms of the nucleus reveal more detail than had formerly been observed. The region from λ 4500 to λ 4700 is especially interesting: strong nuclear, fairly sharp lines of N III (λλ 4634.16 and 4640.64) and of C III (λλ 4647.40 and 4650.16–4651.35; λ 4665.90) are present. Other nuclear lines due to carbon or nitrogen are observed at λ 4267 (C II), λ 4325 (C III), λ 5696 (C III, strong) and λ 4058 (N IV, trace).[11] These bright lines have widths of about 2.35 A (or 155 km./sec.) at λ 4640; their radial velocity is identical with the value obtained for the nebula.

The behavior of He II is similar to that in 9 Sagittae, as is also the absence of O III in emission.[12] All the He II lines, except λ 4686, appear as wide absorptions ($\Delta\lambda$ = 5.7 A at λ 4542 or Δv = 370 km./sec.); λλ 5412, 4542 and 4200 are fairly strong; the other members of the $4f\ ^2F^\circ$–$ng\ ^2G$ series appear as wide absorptions unsymmetrically underlying the bright, sharp Balmer lines. Bright λ 4686 is sharper than the absorption lines; its width is approximately 4.5 A (or 285 km./sec.); its radial velocity is identical with the values derived from the absorption lines and from the nebular emissions. Hence, λ 4686 is emitted in a symmetrical shell while λλ 5412, 4542, 4200, etc., are absorbed in the reversing layer of the underlying star. The excitation of bright λ 4686 must be of similar origin as in

9 Sagittae and in similar objects. It is probably due to the absorption of Lyman α by the He^+-ions in the metastable $2s\ ^2S$-level.

On the other hand, this nucleus differs radically from 9 Sag in that it shows strong carbon lines. The emission is also more marked than in 9 Sag, as compared to the absorption.

We conclude that the nucleus of IC 418 is an O-type star with a strong continuous spectrum upon which are superposed emissions of C III, N III, C II and He II 4686, but which presents the series of $4f\ ^2F^\circ$–$ng\ ^2G$ of He II in absorption. The bright lines are narrower than in pure Wolf-Rayet stars or in 9 Sag; the widths of the N III and C III lines correspond to an ejection velocity of the order of only 80 km./sec. This is the lowest velocity of ejection which we have observed for an object which is similar to Wolf-Rayet stars.[13]

NGC 40.—The nucleus has a typical Wolf-Rayet spectrum, very similar to that of the nucleus of BD 30°3639. It was classified as a *WC*-star, both by Beals and by Payne[14] on the basis of Wright's observations. Our spectrograms show that nitrogen is absent or extremely weak, compared to carbon. The visual region was not described by Wright, but our spectrograms reveal so complete a similarity to BD 30°3639 that a separate description of NGC 40 is not necessary. The nebular spectrum had not been observed in the visual region. Our spectrograms show that the red [N II] lines are present, with a combined intensity similar to that of H_α.

The object is analogous to Campbell's star or to HD 167362 and consists of a pure *WC*8 nucleus surrounded by a nebulosity which is rich in nitrogen.

Other Planetary Nuclei.—A list of tentative classifications of planetary nuclei has been published by Beals. It includes thirteen nuclei, among which nine are considered as *WC* stars and four as *WN* stars. Among the nine nebulae with a carbon nucleus, four had been observed in the red region and all show strong [N II] lines. The present paper shows that one of these nine objects, IC 418, contains both C and N and that its surrounding nebulosity contains also nitrogen. The four nuclei classified by Beals as *WN* objects are: NGC 6543, NGC 6572, NGC 6826 and IC 5217. For NGC 6543 our material shows beyond any doubt[5] that the nucleus contains both C and N; the nebulosity shows also fairly strong [N II] lines despite its high excitation; permitted lines of carbon appear also in the nebula. The spectrum of IC 5217 is not sufficiently well known and requires further observations. For the nuclei of NGC 6572 and NGC 6826, excellent descriptions have been published by Wright. In the nucleus of NGC 6572, the characteristic lines λλ 4634–4641 of N III, λ 4650 of C III, λλ 4659 and 5802–5812 of C IV are present. Hence, the simultaneous presence of C and N is certain. In the nucleus of NGC

6826, Wright observed a broad band extending from λ 4635 to λ 4661; it is presumably a blend of the characteristic N III, C III and C IV lines.

To these objects listed by Beals, we should now add IC 4997, whose nucleus contains both C and N, and HD 167362 which is similar to Campbell's object.

Conclusions.—The planetary nuclei exhibiting bright features belong to two classes:

(*a*) Nuclei of the *WC* type, showing no trace of nitrogen. In all cases where the visual region of the nebular spectrum has been observed, the forbidden lines of [N II] are strong in the surrounding nebulosity. Typical examples are BD 30°3639, NGC 40, HD 167362.

(*b*) Stars showing with similar intensities lines of C and N. In those cases where the visual spectrum has been observed, the nebulosity exhibits [N II]. In one case which has been investigated in detail—NGC 6543—relatively strong permitted lines of C II and C III appear in the nebula. Typical examples are NGC 6543, IC 418, IC 4997, NGC 6572.

To our knowledge no planetary nucleus of Wolf-Rayet type can at present be attributed to the nitrogen sequence. The only exception may be the nucleus of NGC 2392, in which Wright[9] observed bright N III lines and in which we also found weak N V.[12] But this star is essentially an absorption O-type star of the 9 Sagittae type and the emission is not the essential feature.

These characteristics of the planetary nuclei should be compared to those of Wolf-Rayet stars without visible nebulosity. Among the latter there are no intermediate cases between the *WN* and *WC* sequences and the numbers of *WN* and *WC* objects are practically the same. Beals[2] has already called attention to a possible difference between the mean radii of the Wolf-Rayet planetary nuclei and those of the ordinary Wolf-Rayet stars.

In all cases of planetaries which have been observed in the red region, the [N II] lines are present, even if the nucleus is a pure carbon star. This fact is probably of some cosmological importance.

* IC 418 (HD 35914; BD −12°1172); α (1900) $5^h22^m.8$; δ (1900) −12°46′
NG.C 40 (HD 826); α (1900) $T0^h7^m.6$; δ (1900) +71°58′.

[1] *Phys. Rev.*, **55**, 434 (1939).
[2] For example: *Jour. Royal Astr. Soc. Canada*, **34**, 169 (1940)
[3] *Pub. Dom. Ap. Obs., Victoria*, **4**, 286 (1930).
[4] *Zs. f. Ap.*, **7**, 1 (1933).
[5] *Ap. Jour.*, **92**, 289 (1940).
[6] *Ap. Jour.*, **93**, 356 (1941).
[7] *Proc. Nat. Acad. Sci.*, **26**, 548 (1940)
[8] *Proc. Nat. Acad. Sci.*, **26**, 454 (1940).
[9] *Lick Obs. Pub.*, **13**, 193 (1918).
[10] *Lick Obs., Bull.*, **17**, 179 (1935).

[11] There is an unidentified line at λ 4571, also observed in various other high excitation objects. This line probably belongs to the nebula.

[12] *Ap. Jour.*, **91**, 546 (1940).

[13] *Ap. Jour.*, **93**, 354 (1941).

[14] But C. H. Payne includes N III and N IV among the constituents of the spectrum.

RADIOACTIVE CARBON AS AN INDICATOR OF CARBON DIOXIDE UTILIZATION. V. STUDIES ON THE PROPIONIC ACID BACTERIA

BY S. F. CARSON, J. W. FOSTER,* S. RUBEN AND H. A. BARKER

HOPKINS MARINE STATION (OF STANFORD UNIVERSITY), PACIFIC GROVE, CALIFORNIA, THE DEPARTMENT OF CHEMISTRY, AND THE DIVISION OF PLANT NUTRITION, UNIVERSITY OF CALIFORNIA

Communicated April 5, 1941

Previous studies[1,2,3] on CO_2 uptake by the propionic acid bacteria using labeled CO_2 have shown that the two major products of the glycerol fermentation, propionic and succinic acids, contain labeled carbon. The succinic acid has been shown to contain C* only in the carboxyl groups.[2] The results of preliminary experiments using short-lived radioactive carbon (C^{11}) led to the suggestion that the labeled carbon was present in the C_2H_5-group as well as the carboxyl of the propionic acid.[3]

Further experiments have shown this suggestion to be incorrect. The mechanisms of the reactions used to locate the C* in the propionic acid have been investigated. On the basis of the information so obtained the propionic acid formed by the bacteria during a glycerol fermentation in the presence of C^*O_2 has been shown beyond any reasonable doubt to *contain C* only in the carboxyl group.*

The previous experiments on the location of C* in the propionic acid were carried out by means of an alkaline permanganate oxidation. The products are carbonate and oxalate[4] and the reaction follows almost quantitatively the equation

$$\begin{array}{l} CH_3 \\ | \\ CH_2 \\ | \\ COO^- \end{array} + 4MnO_4^- = \begin{array}{l} COO^- \\ | \\ COO^- \end{array} + CO_3^= + 4MnO_2 + OH^- + 2H_2O \qquad (1)$$

The oxidation was effected by alkaline $KMnO_4$, at 100°C. for varying lengths of time with the addition of one gram of inactive propionic acid to act as carrier. The excess MnO_4^- was carefully reduced with H_2O_2

and the MnO_2 removed by centrifugation. The solution was acidified with acetic acid and the liberated CO_2 was absorbed in NaOH solution. The oxalate was then precipitated as the calcium salt, and since the initial radioactivity of the propionate* was known, the extent of decarboxylation could be followed both by gravimetric and radioactivity measurements; the agreement was excellent.

TABLE 1

EXPERIMENTAL CONDITIONS	EXTENT OF DECARBOXYLATION GRAVIMETRIC	EXTENT OF DECARBOXYLATION RADIOACTIVE	C* ACCOUNTED FOR	DISTRIBUTION ON BASIS OF C* RECOVERED OXALATE	DISTRIBUTION ON BASIS OF C* RECOVERED CARBONATE
30 minutes, excess $KMnO_4$ in 15% NaOH	16.5%	17.0%	90%	70%	30%
45 minutes, excess $KMnO_4$ in 25% NaOH	50%	50%	84%	75%	25%

Since the specific activity of the oxalate* fraction was about twice that of the carbonate*, it seemed likely that the C* was present in the alpha and/or beta positions of the propionate*. However, it was recognized that these same results could be obtained from propionate* containing C* only in the carboxyl group since it is not impossible that the carbonate could arise from the ethyl group as well as from the carboxyl carbon.

This latter possibility was tested by direct experiments using long-lived radiocarbon, C^{14}.[5,6] Propionic* acid containing C^{14} only in the carboxyl group was synthesized by the reaction between C_2H_5MgBr and $C^{14}O_2$.[7] The $CH_3CH_2C^*OOH$ was oxidized to oxalate and carbonate by alkaline permanganate at 100°C., and the radioactivity of the oxalate and carbonate fractions was measured with a screen wall Geiger counter. The oxalate fraction contained ~75% of the C*, and the carbonate fraction ~25%, which showed that during the oxidation, fission occurred between the alpha and beta carbons as well as between the alpha and carboxyl carbons. Thus one may conclude that the alkaline permanganate oxidation is not suitable to ascertain the position of the C* in the propionic acid synthesized by the bacteria.

Another reaction that might be satisfactory to locate the C* is the oxidation by means of dichromate in acid solution. It has been reported[8] that propionic acid is oxidized to acetic acid and CO_2 by this reagent. Control experiments carried out with non-radioactive propionic acid in the presence of 0.3 *M* $K_2Cr_2O_7$ in 18 *N* H_2SO_4 showed that the oxidation was complete within three hours at 100°C. and yielded one mole each of CH_3COOH and CO_2 per mole of C_2H_5COOH.

In order to establish the position of the carbon atom from which the CO_2 is derived, radioactive propionic acid ($C_2H_5C^*OOH$) was prepared from $C^{14}O_2$ and C_2H_5MgBr and was oxidized with $Cr_2O_7^=$ under the same

conditions. The resulting acetic acid was completely inactive while all of the C* was found to be located in the C^*O_2. Thus in this oxidation the *CO_2 arises only from the carboxyl carbon.*

Propionic* acid synthesized by the propionic acid bacteria during a glycerol fermentation in the presence of $C^{14}O_2$ was oxidized under the same conditions with $Cr_2O_7^{--}$ and H_2SO_4. The results of this experiment were exactly the same as those obtained with the synthetic $CH_3CH_2C^*OOH$, namely, the acetic acid was completely inactive and all the C* was found in the CO_2 fraction. Hence it may be concluded that the *propionic acid formed by the bacteria contains C* only in the carboxyl group.*

Experiments on the Fermentation of Pyruvate in the Presence of C^*O_2.— In the previous experiments using C*, glycerol was used as the fermentation substrate. It seemed desirable also to study the radioactive organic compounds produced when glycerol is replaced by pyruvate, particularly since with the latter substrate large quantities of acetic as well as propionic acid are formed in accordance with the equation

$$H_2O + 3CH_3COCOOH \rightarrow CH_3CH_2COOH + 2CH_3COOH + 2CO_2$$

A little succinic acid is also produced.

As in previous experiments, the cell suspension technique was used. Thirty-six to 48-hour cultures of *Propionibacterium pentosaceum* were grown in a yeast extract-glycerol medium with phosphate buffer at pH 7.0 under an atmosphere of 5% CO_2 in N_2. Cells from 250 ml. of such cultures were centrifuged from the growth medium and washed twice in phosphate buffer. The washed cells were allowed to shake in buffer for 2 to 3 hours at 30°C. and again washed in order to remove the last traces of glycerol.

For the experiments the cells were shaken anaerobically in 1% phosphate buffer, pH 7.0, and 2% Na pyruvate in the presence of C^*O_2 at 30°C. The fermentation was allowed to proceed for 50 minutes. After removing the bacteria by centrifugation, the volatile and non-volatile products in the supernatant were separated by distillation (inactive propionic, acetic and succinic acids were added as carriers); care was taken to remove all traces of residual C^*O_2.

Radioactivity determinations showed that the volatile acids contained ~5%, the non-volatile materials ~95% of the total activity. This result is to be contrasted with the glycerol fermentation in which the volatile acid fraction contained ~75% of the total radioactivity. A scheme will be discussed below which offers an apparently adequate explanation for this difference.

The C* in the non-volatile fraction could be present in succinate, pyruvate or other as yet unidentified compounds. In view of the high activity in the non-volatile fraction and in view of the possibility that radioactive

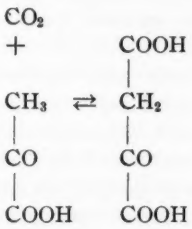
pyruvate could be formed from CO_2 by reversible reactions attention was focused upon the keto-acid fraction. To the non-volatile fraction inactive pyruvic acid was added as carrier, and the 2,4-dinitrophenyl-hydrazone was precipitated. The hydrazone was very active, but, contrary to expectation, only ~10% of this activity was retained after the precipitate was dissolved in Na_2CO_3 solution, washed and reprecipitated several times. It should be particularly mentioned that carrier pyruvic acid was added before each reprecipitation in order to minimize loss of any pyruvate*. If pyruvic acid were the only active keto acid present, very little of the activity should have been lost upon repeated reprecipitation. Therefore it appears that there was a radioactive keto acid other than $CH_3COCOOH$ formed during the fermentation.

A second experiment was performed to further examine this point. The non-volatile fraction was divided into two portions. To one portion were added inactive pyruvic acid and oxaloacetic acid (ethyl ester) as carriers and the 2,4-dinitrophenylhydrazone was precipitated in order to estimate the total amount of active carbonyl compounds. This turned out to be ~70% of the total non-volatile activity. The second portion was acted upon by yeast carboxylase at pH 6.0–6.2, and the CO_2 given off was collected in $Ba(OH)_2$, acid being added to the enzyme solution after the reaction was completed. ~75% of the activity in the non-volatile fraction was liberated as C^*O_2 by the action of the enzyme; the remaining 25% was non-volatile, and very likely due to succinic acid which is known to be produced under these conditions. Since carboxylase is known to be specific for alpha-ketonic acids it seems safe to conclude that a considerable fraction of the non-volatile C* was contained in the carboxyl of an alpha-keto acid.

The identity of the alpha-keto acid has not been established in these experiments. The evidence strongly indicates that a considerable part of the C* is present in some alpha-keto acid other than pyruvic. It is uncertain whether any of the C* is present in pyruvic acid.

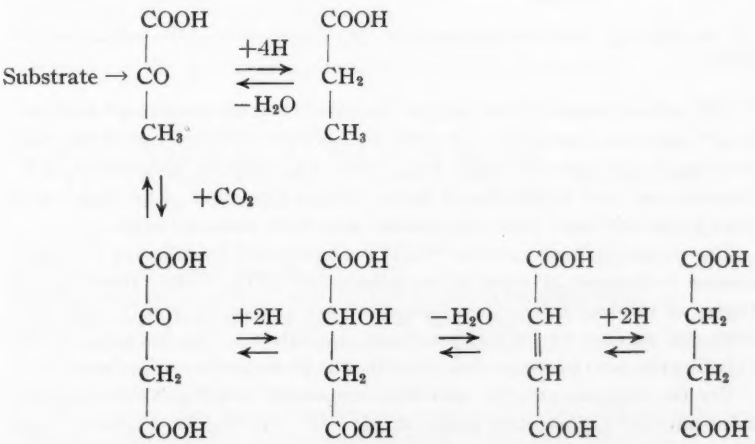
It is of some interest to note that all of the activity extracted (about $^1/_3$ of the total C^*O_2 assimilated) from the cells by vigorous boiling with acid solution was precipitated as an insoluble hydrazone of 2,4-dinitrophenyl-hydrazine. The above results are of considerable interest in connection with the results of other workers. Evans and Slotin[9] have shown that alpha-ketoglutarate* is formed by pigeon liver from bicarbonate*. They further state that "Since pigeon liver can synthesize alpha-ketoglutaric acid from oxaloacetic acid and pyruvic acid,[10] the data suggest, as the simplest hypothesis, that CO_2 combines directly with pyruvic acid to yield oxaloacetic acid, the latter then combining with an additional molecule of pyruvate to form alpha-ketoglutarate by the reactions previously de-

scribed." In addition, Krebs[11] has obtained evidence that the primary step in the synthesis of alpha-ketoglutarate by pigeon liver is:



In view of our experiments which indicate the existence of an alpha-keto acid besides pyruvic acid, it is possible that this or a similar reaction takes place in the propionic acid fermentation as well, and that by further reduction succinic acid is formed. This was suggested by Wood and Werkman[2] as the simplest hypothesis for the formation of labeled succinic acid from labeled bicarbonate by the propionic acid bacteria, before there was any evidence for the existence of an alpha-keto acid other than pyruvic acid in this fermentation. Thimann[15] had previously pointed out the possibility that CO_2 may enter into biological systems by reversible decarboxylation reactions.

The following scheme accounts for the formation of propionic* and succinic* acids in the presence of C^*O_2:



The above scheme is in agreement with all of the known facts concerning the propionic acid fermentation, namely:

1. Both propionic and succinic acids formed in the presence of C^*O_2 are radioactive.

2. These acids contain the labeled carbon only in the carboxyl groups.

3. Pyruvic acid is an intermediate compound[13] in the propionic acid fermentation.

4. A radioactive alpha-keto acid, besides pyruvic acid, is formed during the fermentation of pyruvate in the presence of C^*O_2. This acid contains most, if not all, of the C* in carboxyl groups.

5. The set of reversible reactions from oxaloacetate to succinate have been found to occur in the propionic acid bacteria.[14]

Since fumarate and succinate are symmetrical molecules, the reversible reactions will give rise to oxaloacetic acid containing C* in both carboxyl groups. The decarboxylation of the latter compound will give rise to pyruvic* acid containing C* in the carboxyl group. By reduction, the pyruvic* acid thus formed can yield propionic* acid containing C* only in the carboxyl group.

TABLE 2

INTERCONVERSION OF PROPIONIC AND SUCCINIC ACIDS

SUBSTRATE (INACTIVE)	RADIOACTIVE ACID ADDED	TIME OF FERMENTATION (MINUTES)	PROPIONIC* ACID	RECOVERED[a] SUCCINIC* ACID	C^*O_2
None	Propionic	53	100 ± 10	0 ± 10	3.5 ± 1
Glycerol	Propionic	42	....	5 ± 2	0 ± 2
	Succinic	32	2 ± 2	.....	0 ± 2
Pyruvate	Propionic	100	90 ± 10	9 ± 2	8 ± 1
	Succinic	51	1 ± 1	81 ± 10	0 ± 1

[a] The figures are represented in terms of 100 arbitrary units of the radioactive acid added.

The scheme requires that all the reactions between propionic* and succinic* acids are reversible. In order to ascertain whether this is the case, propionic* and succinic* acids were added separately to suspensions of *P. pentosaceum*, and a search was made for the formation of succinic* acid from propionic* acid, and of propionic* acid from succinic* acid.

The propionic* and succinic* acids were prepared by allowing *P. pentosaceum* to ferment glycerol in the presence of C^*O_2. After removing all traces of residual C^*O_2, the radioactive acids were rigorously separated from one another by exhaustive steam distillation. The bacteria for the experiments were grown as described for the pyruvate investigations.

For the experiments, the cells were suspended in 0.5% $NaHCO_3$ (inactive) plus 0.5% phosphate buffer at pH 7.0. Ten mg. of propionic* and 10 mg. of succinic* acids were introduced into separate vessels in the absence of any substrate, as well as in the presence of 2% Na pyruvate[12] and of 2% glycerol. The gas phase was N_2 and the vessels were shaken at 30°C.

for 25–100 minutes. The bacteria were then removed by centrifugation; from the supernatant liquid the volatile and non-volatile acids, as well as the CO_2, were separated and their radioactivities measured. The results are given in table 2.

These results would seem to bear out the suggestion that both propionic* and succinic* acids may be formed by means of reversible reactions. It should be explicitly stated that no quantitative conclusions regarding the interconversion rates can be drawn from the above results. The first interpretation of the results in table 2 would be that the rate of formation of propionate* from succinate* is too slow to account for the radioactivity usually found in the propionic acid formed in the glycerol fermentation. However in the interconversion experiments the slowest (i.e., the rate determining) step may be the entrance of the succinic acid into the cell. Succinic acid once inside the cell may well be involved in rapid reversible reactions which could account for the formation of radioactive propionic acid.

The above scheme also offers an explanation for the lower relative activity of the propionic acid formed in the pyruvate fermentation as compared to the glycerol fermentation. With pyruvate as the substrate the high concentration of this substance would tend to favor the formation of

$$COOH—CO—CH_2—COOH$$

and at the same time this would decrease the specific activity of the pyruvic acid from which the propionic acid is formed by reduction.

It is a pleasure to express our thanks to Professor C. B. van Niel for much valuable advice and discussion.

We wish to thank Professor E. O. Lawrence, Dr. M. D. Kamen and members of the Radiation Laboratory for their coöperation.

* National Research Council Fellow, 1939–1940. Present address: Research and Development, Merck and Co., Rahway, N. J.

[1] Carson and Ruben, *Proc. Nat. Acad. Sci.*, **26,** 422 (1940).

[2] Wood, Werkman, Hemingway and Nier, *Jour. Biol. Chem.*, **135,** 789 (1940).

[3] Carson, Foster, Ruben and Kamen, *Science*, **92,** 433 (1940).

[4] Przewalski, *Jour. Prakt. Chem.*, (2), **88,** 145 (1913).

[5] Ruben and Kamen, *Phys. Rev.*, **57,** 549 (1940).

[6] Ruben and Kamen, *Ibid.*, **59,** 349 (1941).

[7] The experiments with C^{14} were carried out by Ruben and Nahinsky.

[8] Polonovski and Lindenberg, *Compt. Rend.*, **209,** 46 (1939).

[9] Evans and Slotin, *Jour. Biol. Chem.*, **136,** 301 (1940).

[10] Evans, *Biochem. Jour.*, **34,** 829 (1940).

[11] Krebs and Eggleston, *Ibid.*, **34,** 1383 (1940).

[12] The cells for these pyruvate experiments were freed of residual glycerol in the same manner as described for the first experiments on the fermentation of pyruvate.

[13] Wood and Werkman, *Biochem. Jour.*, **28,** 745 (1934).

[14] Krebs, Private communication.

[15] Thimann, *Science*, **88,** 506 (1938).

RELATION BETWEEN GENES AND CAROTENOIDS OF THE TOMATO

BY A. L. LEROSEN, F. W. WENT AND L. ZECHMEISTER

GATES AND CRELLIN LABORATORIES OF CHEMISTRY* AND THE KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated March 26, 1941

An explanation of the action of Mendelean characters or genes in the development of an organism will have to be based on the chemical changes they produce. But very little is known about the relation between genes and the chemical composition of a plant. This is chiefly due to the difficulties inherent in the genetic analysis of chemical composition, which will require thousands of individual analyses. In one instance, however, chemical differences can easily be recognized by the geneticist, viz., when the genetics of colored substances is studied. In the case of anthocyan colors an extensive study by English investigators has shown that single genes affect such biochemically simple reactions as oxidation, methoxylation and methylation, as well as pigment production or changes in the pH of the cell sap. (See the review by Scott-Moncrieff.[1])

The chromatographic method invented by Tswett is available for the differentiation and quantitative estimation of carotenoids, and has been better developed for these than for any other class of natural pigments[2]. Since a number of genes affecting carotenoids are known, it is surprising that not more work on the chemistry of carotenoid inheritance has been done. For lower plants the contribution of Kuhn and Moewus,[3] relating the crocin, pikrocrocin and safranal content of Chlamydomonas to its genetic behavior, stands out. In the higher plants we can quote articles by Mangelsdorf and Fraps[4] as well as Randolph and Hand[5] who have shown that the provitamin A content of corn endosperm is a function of the number of *Y* genes. Work by Emsweller, Burrell and Borthwick[6] and by Johnson and Miller[7] on inheritance of carotenoids in carrots and wheat was carried out by the spectroscopic analysis of extracts, which makes identification of some individual carotenoids more difficult. This is also true of work described in a short article by Brass, Beyrodt and Mattausch,[8] who tentatively identify the yellow pigment of tomato skin as β-carotene without mentioning genetic relations.

The present paper deals with the qualitative and quantitative estimation of the components of the polyenic pigment of ripe tomatoes as influenced by genes. Two sets of genes are known to affect the color of the ripe tomato fruit (Lindstrom[9]). Of the skin color genes, *Y* and *y*, the dominant form *Y* causes a yellow color of the epidermis, the homozygous recessive *y* resulting in a colorless skin. Another pair of genes produces

red, *R*, or yellow, *r*, flesh. The genes mentioned are responsible for the well-known difference between the red and yellow tomato. The two sets of genes are completely independent and are presumably located in different chromosomes.

For the present investigation four sets of ripe tomato fruit were kindly supplied by Dr. J. Leslie of the Citrus Experiment Station in Riverside, California. They were all collected in the same field at the same time. Within each group various plants with differing genetic constitutions were represented, but in size and color the fruit were uniform. The genetic constitution of the fruit was:

Yr: yellow skin, yellow flesh (golden yellow tomato),

yr: colorless skin, yellow flesh (paler than the previous one),

yR: colorless skin, red flesh (a red tomato lacking the warm orange tinge),

YR: yellow skin, red flesh (common red tomato).

The skin was peeled off and scraped as clean as possible from adherent flesh. Microscopic examinations showed, however, that a certain number of flesh cells adhered to the epidermis. The "colorless skin" tomatoes, *y*, had no recognizable color in their epidermis cell walls, but the "yellow skin" fruits, *Y*, had bright yellow epidermal walls. Based on the microscopic inspection of the peeled skins it is estimated that in Nos. 1–4 (table 1), at least 10% of their volume was composed of flesh cells and their remains (for No. 4*a* see later).

By means of the chromatographic method, described below, the amounts of the individual pigments have been determined, as given in table 1.

TABLE 1

INFLUENCE OF GENES ON THE CAROTENOID CONTENT OF CERTAIN RIPE TOMATOES

(The figures indicate milligrams in 100 grams of dried material, with the exception of the alkali-soluble yellow pigment, which is expressed in arbitrary units.)

NO.	GENE	MATERIAL	XANTHOPHYLLS	RED CAROTENOIDS*	YELLOW HYDROCARBONS	ALKALI-SOLUBLE YELLOW PIGMENT
1	*Yr*	Skin†	0.7	0	0.3	18
2	*yr*	"	0.6	0	0.4	2
3	*yR*	"	1.9	35.5	5.9	2
4	*YR*	"	2.6	131	5.8	20
4*a*	*YR*	"	..	61‡	..	..
5	*Yr*	Flesh	3.6	0	3.4	1
6	*yr*	"	4.0	0	2.4	1
7	*yR*	"	6.4	229	24.0	2
8	*YR*	"	10.4	216	15.0	2

* These figures include small amounts of γ-carotene and some related hydrocarbons (details in table 2).

† With small amounts of flesh attached (see text).

‡ In this case the skins have been freed from flesh with special care and under microscopic control, until practically no more flesh cells or debris were observed.

Discussion.—The yellow color of the skin of the tomatoes investigated is due to an as yet unidentified alkali-soluble pigment already mentioned by other workers.[10] This pigment is located in the thick cell walls of the epidermis. Small amounts were obtained from the flesh; these amounts were independent of the genetic constitution of the plant (table 1, Nos. 5–8). A similar quantity was found in the colorless skins (Nos. 2 and 3, *yr* and *yR*). This indicates that at least a small, almost constant amount of the alkali-soluble yellow pigment is present everywhere in the four groups of tomato fruits. The gene *Y* in dominant form increases this amount tenfold (Nos. 1 and 4) in the skin but leaves the flesh unaffected (Nos. 5 and 8 compared with 6 and 7). Whether the dominant gene *R* or its recessive modification *r* is present does not change the amount of the alkali-soluble pigment significantly.

The gene *R* is clearly responsible for the increase in carotenoids proper both in flesh and skin. When present in dominant form, it determines in the first place the presence of lycopene (Nos. 3–4 and 7–8), but the increase in this main pigment is always accompanied by increases in the other carotenoids; the yellow hydrocarbons were increased eightfold, the xanthophylls 2 5 times. It is an open question whether we have to deal with a pleiotropic effect of the gene *R*, or whether the parallel increases in all carotenoids are due to a series of linked genes. This can be determined only by future genetic analysis, in which the basis of classification will be chromatograms instead of the gross color of the fruit.

One point needs further discussion, viz., the variability of the carotenoid content in the skin. This is attributable to different causes. In the first place, there are considerable differences in the chromoplast content of the epidermal cells. Whereas in some instances a relatively large number of lycopene crystals can be seen in the epidermal cells, in other cases the latter seem relatively free from plastids (cf. Kylin[11] who found only very small chromoplasts in the epidermis of tomato fruits). In the second place, varying amounts of flesh cells adhered to the scraped epidermis. In *one* sample (4*a*), however, which was prepared with special care, microscopic examination failed to show any appreciable amount of adhering flesh cells; nevertheless the lycopene content was of the same order of magnitude as in the other two samples (Nos. 3–4). Thirdly, variations in the thickness of the epidermal walls obviously affect the percentage of carotenoids in the skin, since by far the largest part of the dry weight of the skin is made up of cell walls.

Experimental.—Starting materials: Each portion (2.65 kilos = 58–68 tomatoes) was peeled and the skins were kept in water overnight to facilitate the subsequent removal of flesh particles. The wet skins were scraped individually with a knife, and dehydrated in methyl alcohol for 12 hours. The alcohol was then removed on a Büchner funnel. The sample was

dried for 6 hours at 40° (4.5–7 g.) and kept under carbon dioxide. The alcoholic filtrate was diluted with water and extracted with ether, which was washed free of alcohol, dried over sodium sulfate and used in the subsequent extraction of the skin.

The pulp (free from seeds, about 0.8 kilo) was washed, cut up in a meat grinder, sucked as dry as practicable in a Büchner funnel, kept in methanol overnight and treated as described above for the skins. This operation yielded a dried pulp (10.5–12 g.) and an ether extract.

Chromatography: Each sample (dry pulp or skin) was divided into two parts for duplicate estimations. About 2–6 g. were ground in a coffee mill and treated first with the corresponding half of the ether extract (see above) and then repeatedly with peroxide-free ether during two days until the last extract was almost colorless. The solution (500 cc.), when completely evaporated in a slow CO_2 stream, finally in vacuum, left a partially crystalline residue if lycopene was the main pigment; in other cases it was oily. The residue was dissolved in just the necessary amount of petroleum ether† or, if the lycopene content was high, in benzene. In the latter case 3 volumes of petroleum ether were added. The solution was immediately chromatographed on calcium hydroxide (Shell; cylindrical part of tube:‡ 23 × 2.5 cm.; for lycopene rich solutions: 25 × 4 cm.). In the absence of lycopene petroleum ether was used as a developer; if it was present, the chromatogram was developed first with a 1:1 mixture of petroleum ether and benzene, then with pure benzene and finally the latter was displaced with petroleum ether.

If the lycopene content is low all components of the pigment can be retained in the column. In the opposite case lycopene occupies such a large section of the adsorbent that it is advisable to wash the carotene into the filtrate, which is evaporated, dissolved in a small volume of petroleum ether, and rechromatographed on calcium hydroxide. By the latter procedure the following two chromatograms, *A* and *B*, were obtained. (The figures on the left side denote the height of the zones, in millimeters. Adsorption maxima (mμ) refer to petroleum ether solutions.)

Chromatogram A:

3 pale brownish yellow (oxidized pigment)
several yellow and brown lines: xanthophylls 471, 449 (424)
5 very pale yellow diffuse pigment
4 purple tinged pink: lycoxanthin 504, 473, 447
17 nearly colorless
65 red: lycopene 504, 474, 445
35 orange: neo-lycopene 499.5, 468.5, 440
5 light yellow: lycopene isomer 496.5, 466 (439)

7 nearly colorless
7 pink: γ-carotene (?) 492.5, 460, 431
10 almost colorless
10 lemon-yellow: unidentified carotene 490.5, 457, 429.5 (404)

Chromatogram B (as prepared from the filtrate of *A*):

8 lemon-yellow: unidentified carotene 490.5, 457, 429.5 (404)
3 almost colorless
20 orange: unidentified carotene isomer 480.5, 450 (426)
5 almost colorless
28 deep orange: β-carotene 486, 453 (427)
2 almost colorless
25 bright orange: unidentified carotene 488.5, 454
2 almost colorless
25 orange-yellow: α-carotene 478, 446.5

The column was cut in an appropriate way, and the individual layers were eluted with a 3:1 mixture of petroleum ether and alcohol (95%). Each solution was washed free of alcohol and dried with sodium sulfate. The spectra were determined in an Evaluating Grating Spectroscope as devised by Loewe and Schumm (Zeiss; light filter: Jena *BG* 7) and the concentration of the pigments in the Pulfrich Gradation Photometer (Zeiss; light filters *S*45 for β-carotene, *S*47 for the other polyenes; photometric values as given by Cholnoky[12]:). The sum of xanthophylls was determined as "lutein," lycoxanthin as "lycopene," neolycopene and the next lower layer as "lycopene" × 1.5; for γ-carotene the mean photometric values of lycopene and β-carotene were adopted; the next four layers were determined as "β-carotene."§ The figures obtained are given in table 2.

Alkali-soluble yellow pigment: After the carotenoids had been removed with ether the respective residues of the pulp and of the skin were extracted once with 100 cc. of *N* NaOH, twice with 50 cc. of *N*/4 alkali, and finally several times with water (total volume 500 cc.). By using the filter *S*47 in the photometer, an arbitrary unit was chosen, viz., that amount contained in 500 cc. of *N*/4 NaOH, which gave $k = 0.03$, in a 1-cm. cell. (The curve is approximately linear, 10 units showing $k = 0.39$.)

The main source of experimental error in the determination of the carotenoids was incomplete extraction. The material after having been apparently exhausted with ether yielded a new fraction when kept in the solvent for 2 weeks. Such a procedure may involve, however, the risk of autoxidation and partial bleaching. In some instances the error amounted to 12–28% of the total lycopene content. This does not affect the genetical deductions.

TABLE 2

THE CONTENT OF CHROMATOGRAPHICALLY SEPARATED CAROTENOIDS IN CERTAIN TOMATOES (MG. IN 100 G. OF DRY MATERIAL)
("–" means present but not determined)

NO.	GENE	MATERIAL	XANTHOPHYLLS	LYCOXANTHINE	LYCOPENE	NEOLYCOPENE	UNIDENTIFIED LYCOPENE ISOMER	γ-CAROTENE	UNIDENTIFIED CAROTENE	UNIDENTIFIED CAROTENE	β-CAROTENE	UNIDENTIFIED CAROTENE	α-CAROTENE
1	*Yr*	Skin	0.7	0	0	0	0	0	0	–	0.3	0	–
2	*yr*	"	0.6	0	0	0	0	0	0	–	0.4	0	–
3	*yR*	"	1.9	1.9	23.1	8.8	1.4	0.3		1.1	4.8	–	–
4	*YR*	"	2.6	1.3	107.0	22.8	–	0.9		1.5	4.3	–	–
4a*	"	"	–	–	61	–	–	–		...	0.8	–	–
5	*Yr*	Pulp	4.3	0	0	0	0	0	0	1.0	1.8	0	0.6
			2.9	0	0	0	0	0	0	0.9	1.8	0	0.7
6	*yr*	"	3.5	0	0	0	0	0	0	0.6	1.2	0	0.5
			4.5	0	0	0	0	0	0	–	1.9†	0	0.5
7	*yR*	"	6.2	8.8	198	38.0	–	1.7	1.7	–	(24)†	–	–
			6.5	6.6	146	55.5	–	1.6	0.8	3.8	15.0	4.8	1.4
8	*YR*	"	10.2	14.5	190	13.8	–	0.4	0.7	1.8	28.8‡	–	–
			10.5	10.1	178	22.5	1.6	–	0.3	3.3	7.8	0.9	3.0

* Specially prepared with microscopic control (see text).

† These figures include the foregoing and the subsequent layers.

‡ Includes the two subsequent layers.

Summary.—This work has shown that two pairs of genes, *R* and *r*, causing red or yellow flesh color in tomatoes, and *Y* and *y*, producing yellow or colorless skin, are completely unrelated not only genetically but also in their chemical effect. The gene *R* affects only the red and yellow plastid pigments: chiefly lycopene, to a much less extent carotenes and xanthophylls. The gene *Y* has no influence on these plastid pigments, but causes a tenfold increase in an alkali-soluble yellow pigment which generally occurs in tomatoes. This latter effect is restricted to the epidermis.

* Contribution No. 817.

† The petroleum ether referred to throughout this paper is that of b. p. 60–70°.

‡ A photograph of this device can be found in Ref. (2), p. 62, figure 19. The tubes are now manufactured by Scientific Glass Apparatus Co., 49 Ackerman St., Bloomfield, N. J.

§ On saponification all unidentified layers kept their epi-phasic behavior in the partition test.

[1] Scott-Moncrieff, R., "The Genetics and Biochemistry of Flower Colour Variation," *Ergebn. Enzymforsch.*, **8**, 277 (1939).

[2] Zechmeister, L., and Cholnoky, L., *Die chromatographische Adsorptionmethode*, 2nd ed., Vienna, Springer (1938), English translation in print (Chapman & Hall, London).

[3] Kuhn, R., and Moewus, F., *Ber. chem. Ges.*, **73**, 547 (1940).

[4] Mangelsdorf, P. C., and Fraps, G. S., *Science*, **73**, 241 (1931).

[5] Randolph, L. F., and Hand, D. B., *Jour. Agric. Res.*, **60**, 51 (1940).

[6] Emsweller, S. L., Burrell, P. C., and Borthwick, H. A., *Proc. Amer. Soc. Hort. Sc.*, **33**, 508 (1936).

[7] Johnson, I. J., and Miller, E. S., *Cereal Chem.*, **15**, 345 (1938).

[8] Brass, K., Beyrodt, A., and Mattausch, J., *Naturwissensch.*, **25**, 60 (1937).

[9] Lindstrom, E. W., *Genetics*, **10**, 305 (1925).

[10] Smith, O., *Cornell Univ. Agric. Exp. Stat. Mem.*, **187**, 3 (1936); Euler, H. von, Karrer, P., Krauss, E. von, and Walker, O., *Helv. Chim. Acta*, **14**, 154 (1931).

[11] Kylin, H., *Z. physiol. Chem.*, **163**, 229 (1927).

[12] Cholnoky, L., *Untersuch. Lebensmittel* (in print).

ON THE FIRST VARIATION OF THE DIRICHLET-DOUGLAS INTEGRAL AND ON THE METHOD OF GRADIENTS

By Richard Courant

New York University

Communicated April 7, 1941

Euler's equations and the natural boundary conditions in a variational problem are a consequence of the vanishing of the first variation of the functional for a properly selected class of variations of the independent functions. This class may be chosen in different ways, and thereby essential differences in the treatment may originate, as shown by the Plateau-Douglas problem for minimal surfaces. Douglas[1] (who intro-

duced the Dirichlet integral as the basis of the theory) and, for higher type-numbers of stationary solutions, Morse and Tompkins [5] operate in a restricted space $\mathfrak{P}$ of harmonic vectors $\mathfrak{x}$, while the author of the present note [1, 2, 3] and M. Shiffman [6] preserve the freedom of operating in a wider space $\mathfrak{S}$, a procedure anticipated by T. Radò for simply connected surfaces. As pointed out by Morse [4], the equivalence of these different approaches does not seem obvious if one wants to characterize a minimal surface by the vanishing of the variation of the Dirichlet functional for those classes of independent variations in $\mathfrak{P}$ and $\mathfrak{S}$ which are actually used in [1, 2] and [3] or [4, 5], respectively.

The first remark in this note establishes this equivalence (§ 2) and provides the link between the theory in [4] and that in [1] or [2] (§ 3).

Our second remark (§ 4) refers to the first variation in the space $\mathfrak{P}$ of harmonic vectors $\mathfrak{x}$. In 1908 Hadamard [7] suggested a general method for a constructive existence proof in variational theory. While this "method of gradient" or "method of steep descent" has met with difficulties in problems in more dimensions with fixed boundary values, the Plateau-Douglas problem and that of conformal mapping of simply or multiply connected domains furnish the first instance where it can be successfully used for the construction of the solution.

For both purposes we shall rely on the same general expression of the first variation by a double integral whose integrand is a divergence expression if the vectors $\mathfrak{x}$ are harmonic.

1. Definitions. Variational Formula.—We consider a plane domain B with the analytic boundary C, variable within a certain class (e.g., k-fold connected domains bounded by circles). The domains may lie in the plane of the complex variable $w = u + iv$ or $w' = u' + iv'$. Admissible vectors, forming the space $\mathfrak{S}$, are vector functions $\mathfrak{x}(u, v)$ continuous in $B + C$ with piecewise continuous first derivatives in B, with a finite Dirichlet integral

$$D(\mathfrak{x}) = {}^1/_2 \int\int_B (\mathfrak{x}_u^2 + \mathfrak{x}_v^2) du dv,$$

and such that they map C in a continuous and monotone way on a prescribed system Γ of Jordan curves in the $\mathfrak{x}$-space. With respect to this class $\mathfrak{S}$ of functions and the class of domains B we define a type of variations $\mathfrak{A}$ of the independent variable $\mathfrak{x}$ and of the functional $D(\mathfrak{x})$ as follows: We map B by a one-one transformation onto an admissible domain B' in the w' plane:

$$u = u' + \epsilon\Lambda, \quad v = v' + \epsilon\mathrm{M}$$

where the quantities Λ and M are considered as functions of u, v, ϵ or of u', v', ϵ; for sufficiently small $|\epsilon|$ and for u, v in $B + C$ they may have

piecewise continuous first derivatives whose absolute values are bounded by a fixed number a. For $\epsilon = 0$ the quantities Λ and $\mathbf{M}$ are called λ and μ. The variation $\mathfrak{A}$ of the vector $\mathfrak{x}$ consists in replacing $\mathfrak{x}$ in B by a vector $\mathfrak{z}$ in B' defined by

$$\mathfrak{z}(u', v') = \mathfrak{x}(u, v).$$

Referring to the domain B' here and later by a prime we calculate the Dirichlet integral $D'(\mathfrak{z})$ by substituting $\mathfrak{z}_{u'} = \mathfrak{x}_u(1 + \epsilon\Lambda_u) + \mathfrak{x}_v\epsilon\mathbf{M}_u$, etc., rearranging terms, and applying Schwarz's inequality. Thus we obtain the formula[2]

$$D'(\mathfrak{z}) = D(\mathfrak{x}) + \epsilon V(\mathfrak{x}, \lambda, \mu) + \epsilon^2 R, \tag{1}$$

where

$$|R| < cD(\mathfrak{x}),$$

(c depending only on the bound a for the derivatives of Λ and $\mathbf{M}$ and on a bound for ϵ), and where the first variation V is defined by

$$V(\mathfrak{x}, \lambda, \mu) = {}^1/_2 \int_B \int [p(\lambda_u - \mu_v) - q(\lambda_v + \mu_u)]dudv \tag{2}$$

with

$$p = \mathfrak{x}_u{}^2 - \mathfrak{x}_v{}^2, \quad q = -2\mathfrak{x}_u\mathfrak{x}_v.$$

The formula (1) and (2) is general with respect to the vector $\mathfrak{x}$ in $\mathfrak{S}$. However, we shall assume from the outset that the initial vector $\mathfrak{x}$ satisfies the Euler condition $\Delta\mathfrak{x} = 0$, i.e., is harmonic, while of course the varied vector $\mathfrak{z}$ in general will not remain harmonic.[3] Thus $\mathfrak{x}$ is in $\mathfrak{P}$, while $\mathfrak{z}$ may be in the wider space $\mathfrak{S}$. The relation $\Delta\mathfrak{x} = 0$ implies that

$$\varphi(w) = p + iq$$

is analytic in w.

2. Connection between Variations in $\mathfrak{S}$ and in $\mathfrak{P}$.—If we admit only vectors in $\mathfrak{P}$, then $\mathfrak{z}$ is not admissible. But we obtain a variation in $\mathfrak{P}$, uniquely defined by $\mathfrak{z}$, if we replace $\mathfrak{z}$ by a vector $\mathfrak{h}$ harmonic in B' and having the same boundary values as $\mathfrak{z}$. The transition from $\mathfrak{x}$ to $\mathfrak{h}$ is called the variational process $\mathfrak{B}$; it proceeds in the space $\mathfrak{P}$. Now we state the theorem:

In the variational process $\mathfrak{B}$ the first variation of the Dirichlet integral

$$\lim_{\epsilon \to 0} \frac{D'(\mathfrak{h}) - D(\mathfrak{x})}{\epsilon} = S(\mathfrak{x}, \lambda, \mu)$$

exists and is identical with the first variation $V(\mathfrak{x}, \lambda, \mu)$ obtained by the process $\mathfrak{A}$ in the wider space $\mathfrak{S}$.

It should again be emphasized that this theorem refers to initial vectors $\mathfrak{x}$ in $\mathfrak{P}$ and that the difference in the variational processes $\mathfrak{A}$ and $\mathfrak{B}$ is contained in the distinction between the vectors $\mathfrak{z}$ and $\mathfrak{y}$ only.

Proof.—By Dirichlet's principle we have

$$D'(\mathfrak{y}) \leq D'(\mathfrak{z}), \tag{3}$$

so that because of (1)

$$D'(\mathfrak{y}) \leq D(\mathfrak{x}) + \epsilon V(\mathfrak{x}, \lambda, \mu) + \epsilon^2 R. \tag{4}$$

This incidentally shows that $D'(\mathfrak{y})$ is uniformly bounded in ϵ.

In the same way, instead of varying $\mathfrak{x}$ by $\mathfrak{z}$ into $\mathfrak{y}$, we perform the variation in the opposite direction, starting from $\mathfrak{y}$ in B', defining in B the vector $\mathfrak{y}(u, v) = \mathfrak{y}(u', v')$, and finally replacing $\mathfrak{y}$ by the harmonic vector with the same boundary values, that is, by $\mathfrak{x}$. In the transition from B' to B the quantities $-\Lambda$ and $-\mathbf{M}$ play the previous rôle of Λ and $\mathbf{M}$. With $p' = \mathfrak{y}_u^2 - \mathfrak{y}_v^2$ and $q' = -2\mathfrak{y}_u\mathfrak{y}_v$ we define

$$V'(\mathfrak{y}, \lambda, \mu) = {}^1/_2 \int\int_B [p'(\lambda_{u'} - \mu_{v'}) - q'(\lambda_{v'} + \mu_{u'})] du'dv'. \tag{2'}$$

The uniform boundedness of $D'(\mathfrak{y})$ guarantees the existence and uniform boundedness of $|V'|$ as well as the uniform boundedness of $|R'| < cD'(\mathfrak{y})$ in the formula

$$D(\mathfrak{x}) \leq D'(\mathfrak{y}) - \epsilon V'(\mathfrak{y}, \lambda, \mu) + \epsilon^2 R', \tag{4'}$$

which is obtained like (4) in the opposite variational process. We have for $\epsilon \to 0$

$$V'(\mathfrak{y}, \lambda, \mu) \to V(\mathfrak{x}, \lambda, \mu), \tag{5}$$

as will be shown presently. Then we obtain from (4) and (4') after dividing by ϵ (which we take positive),

$$-\epsilon R' + V'(\mathfrak{y}, \lambda, \mu) \leq \frac{D'(\mathfrak{y}) - D(\mathfrak{x})}{\epsilon} \leq \epsilon R + V(\mathfrak{x}, \lambda, \mu)$$

and therefore because of (5)

$$\lim_{\epsilon \to 0} \frac{D'(\mathfrak{y}) - D(\mathfrak{x})}{\epsilon} = V(\mathfrak{x}, \lambda, \mu). \tag{6}$$

This is the statement of our theorem.

To justify (5) we divide B into a closed interior domain B_1 and a boundary strip B_2 with width less than a small parameter δ. A corresponding division of B' is expressed by $B' = B_1' + B_2'$, etc. By well-known theorems of potential theory we have, for $\epsilon \to 0$, the relation $\mathfrak{y} \to \mathfrak{x}$; and the convergence of $\mathfrak{y}$ and of its derivatives is uniform in B_1. Hence in the

integral V_1' referring to B_1' the passage to the limit may be performed under the integral sign; thus we obtain $V_1' \to V_1$. From (4) and (4') it follows immediately that $D'(\mathfrak{h}) \to D'(\mathfrak{x})$. By choosing δ sufficiently small we can make $D_2(\mathfrak{x})$ arbitrarily small, therefore also $D_2'(\mathfrak{h})$ uniformly small. The expression for V_2' then shows, because of the boundedness of the derivatives of λ and μ, that $| V_2 |$ remains uniformly and arbitrarily small if the width δ is chosen properly. This proves (5).

3. Relation to Results of Morse.—If $\mathfrak{x}$ is a minimal surface, we have $p = q = 0$ and hence $V = S = 0$ for all admissible variations in $\mathfrak{P}$ or $\mathfrak{S}$. To infer conversely from $V = S = O$ the relation $\varphi(w) = p + iq = 0$, characterizing $\mathfrak{x}$ as a minimal surface, it suffices to exploit only a narrow subclass of admissible variations. We assume in this section that B is a plane domain bounded by k circles. In [1, 2] it is shown that the characteristic relation $\varphi(w) = 0$ is a consequence of the following conditions $\mathfrak{R}$: (1) the function $(w - w_0)^2\varphi(w)$ is real on the boundary circle with center w_0; (2) the integrals $\int \varphi(w)dw$, $\int w\varphi(w)dw$ vanish for any closed contour in B. The same conditions $\mathfrak{R}$ are used in [4] for concluding $\varphi(w) = 0$. In [4] it is stated that the conditions $\mathfrak{R}$ and hence $\varphi(w) = 0$ follow if $S = 0$ is established for arbitrary variations of the circular domain and for a subclass of variations of the boundary values which may be expressed in polar coordinates r, θ for a boundary circle of B by $w' = we^{i\alpha}$, where on the boundary α is analytic and otherwise arbitrary in θ. On the other hand, in [1] or [2] the conditions $\mathfrak{R}$ are established as a consequence of $V = 0$ for variations with the same boundary values, α being a Poisson kernel. Hence by the theorem of §2 the result of [4] follows from the results in [1] or [2] obtained in the wider class $\mathfrak{S}$.

Morse's method of establishing $\mathfrak{R}$ is based on the continuous dependence of the variation S on its arguments, a result of importance in itself. It may be remarked that the form (2) of the variation V displays such continuity properties to an extent sufficient for the proof of $\mathfrak{R}$, though under conditions apparently more restrictive than formulated in [4]. If the variation α of the boundary values is analytic, then our functions λ and μ may, in strips adjacent to the boundary, be chosen as conjugate harmonic; thereby V is reduced to an integral over a closed interior subdomain of B. As a consequence, for such variations of $\mathfrak{x}$ in $\mathfrak{P}$ the variation V of the functional depends continuously on the initial vector $\mathfrak{x}$; for $\mathfrak{x}$ and its derivatives, hence p and q, vary continuously with $\mathfrak{x}$ in a closed subdomain of B if the boundary values of the harmonic vector $\mathfrak{x}$ on C or the boundary C itself vary continuously, α remaining fixed. Furthermore, as emphasized by Morse in [4], it becomes apparent that S retains its meaning even if $D(\mathfrak{x})$ should be infinite.

4. Remark on the Method of Steepest Descent.—The first variation of a functional can be interpreted as the inner product of its gradient with the

vector representing the variation in the function space. This suggests the following analogy: In an Euclidean x-space we consider the lines of steepest descent of a non-negative differentiable function $f(x)$. These lines, starting from arbitrary points, must lead to stationary points where the gradient vanishes. Similarly, in a function space with a non-negative functional one might try to obtain stationary points by starting with some arbitrary admissible function in the space and by performing, continuously or stepwise, such variations as are indicated by the gradient; thereby the value of the functional decreases at a suitably fast rate as long as the gradient, in some measure, differs noticeably from zero. We expect that such a process should lead to stationary points in the function space.

In the minimal surface theory this general idea can be successfully applied in various forms and for various purposes. Here we confine ourselves to a brief discussion of the simplest Plateau problem where the existence proof appears under a new light. B is now the unit circle, and we may impose the usual three-point condition on the harmonic vectors of $\mathfrak{P}$. As measure δ for the deviation of a vector $\mathfrak{x}$ of $\mathfrak{P}$ from minimal surface character (which also measures deviation of the gradient of $D(\mathfrak{x})$ from zero) we may choose $\delta = \max \left| \Im(w^2\varphi(w)) \right|$, where $|w| < 1/2$. Then our formula for the first variation, if for α a suitable Poisson kernel is substituted, shows readily that $D(\mathfrak{x})$ can be decreased by a quantity proportional to δ with a fixed factor of proportionality.[4] Repeating the process indefinitely we obtain a sequence of vectors $\mathfrak{x}_1, \mathfrak{x}_2, \ldots$ in $\mathfrak{P}$ and corresponding deviations $\delta_1, \delta_2, \ldots$, starting from an arbitrary $\mathfrak{x}$ in $\mathfrak{P}$. Because of the compactness of the class $\mathfrak{P}$, the sequence $\mathfrak{x}_n$ defines a limiting element, again called $\mathfrak{x}$, for which the deviation δ must vanish since δ depends continuously on $\mathfrak{x}$ and since obviously $\delta_n \to 0$. This shows that the process results in the construction of a minimal surface for which $D(\mathfrak{x})$ is stationary (not necessarily a minimum).

The procedure is flexible and capable of generalizations. For example, other variations could be used such as the discontinuous variations introduced in [1, 2],[5] and the method can be applied to the general Douglas problem. It does not give a classification of the stationary points according to type-numbers. But, as will be shown elsewhere, it can be used for an analysis of problems such as the continuous dependence of stationary vectors $\mathfrak{x}$ on the boundary curves Γ.

[1] A detailed reference to Douglas' papers is found in [1]. Numbers in brackets refer to the bibliography below.

[2] See the detailed calculation in [1], p. 61.

[3] One might consider more general variations by adding to $\mathfrak{x}$ a vector with vanishing boundary values. This, however, would be useless since the corresponding variation of $D(\mathfrak{x})$ for harmonic $\mathfrak{x}$ vanishes.

[4] See [6], p. 850.

[5] These variations are essentially used in a paper by Max Shiffman on non-relative minima submitted to the *Annals of Mathematics*.

[1] Courant, R., "The Existence of Minimal Surfaces of Given Topological Structure under Prescribed Boundary Conditions," *Acta Mathematica*, **72**, 51–90 (1940).

[2] Courant, R., "Plateau's Problem and Dirichlet's Principle," *Ann. Math.*, **38**, 679–724 (1937).

[3] Courant, R., "Critical Points and Unstable Minimal Surfaces," these PROCEEDINGS, **27** (1), 51–57 (1941).

[4] Morse, Marston, "The First Variation in Minimal Surface Theory," *Duke Journal*, **6**, 263–289 (1940).

[5] Morse, Marston, and Tompkins, C. B., "The Existence of Minimal Surfaces of General Critical Types," *Ann. Math.*, **40**, 443–472 (1939).

[6] Shiffman, Max, "The Plateau Problem for Non-relative Minima," *Ann. Math.*, **40**, 834–854 (1939).

[7] Hadamard, J., "Mémoire sur le probleme d'analyse relatif à l'équilibre des plaques elastiques encastrées," *Mémoires presentés par divers savants étrangers a l'Académie des Sciences de l'Institut de France*, (2), **33**, 4 (1908).

[8] Lewy, Hans, "Ueber einen Ansatz zur numerischen Lösung von Randwertproblemen," Göttinger Nachrichten, 1925, pp. 118–121.

[9] Courant, R., "Bemerkungen zur Frage der numerischen Auflösung von Randwertproblemen, die aus der Variationsrechnung entspringen," Göttinger Nachrichten, 1925, pp. 122–127.